ELSEVIER

Remote Sensing of Environment 84 (2003) 342–349

Remote Sensing of Environment

www.elsevier.com/locate/rse

Effects of landscape characteristics on land-cover class accuracy

Jonathan H. Smith[a,*], Stephen V. Stehman[b,1], James D. Wickham[c,2], Limin Yang[d,3]

[a]US Environmental Protection Agency, Landscape Characterization Branch E243-05, Research Triangle Park, NC 27711, USA
[b]Department of Forestry, SUNY College of Environmental Science and Forestry, Syracuse, NY 13210, USA
[c]US Environmental Protection Agency, Landscape Characterization Branch E243-05, Research Triangle Park, NC 27711, USA
[d]Raytheon-ITSS, Eros Data Center, Sioux Falls, SD 57198, USA

Received 2 April 2002; received in revised form 1 August 2002; accepted 2 August 2002

Abstract

The effects of patch size and land-cover heterogeneity on classification accuracy were evaluated using reference data collected for the National Land-Cover Data (NLCD) set accuracy assessment. Logistic regression models quantified the relationship between classification accuracy and these landscape variables for each land-cover class at both the Anderson Levels I and II classification schemes employed in the NLCD. The general relationships were consistent, with the odds of correctly classifying a pixel increasing as patch size increased and decreasing as heterogeneity increased. Specific characteristics of these relationships, however, showed considerable diversity among the various classes. Odds ratios are reported to document these relationships. Interaction between the two landscape variables was not a significant influence on classification accuracy, indicating that the effect of heterogeneity was not impacted by the sample being in a small or large patch. Landscape variables remained significant predictors of class-specific accuracy even when adjusted for regional differences in the mapping and assessment processes or landscape characteristics. The land-cover class-specific analyses provide insight into sources of classification error and a capacity for predicting error based on a pixel's mapped land-cover class, patch size and surrounding land-cover heterogeneity.

1. Introduction

Maps are increasingly being used to describe the spatial distribution and pattern of land cover and the impact of human decisions on the landscape. Such maps rely on data gathered by the large number of sensors, both aerial and satellite, that acquire images of the Earth's surface. These images are converted to land-cover maps through the utilization of any number of classification algorithms that link recorded digital number values, or some derivative, to specific land-cover classes (Jensen, 1996, 2000). Effectively employing these land-cover maps requires evaluating their accuracy and presenting results to users (Congalton & Green, 1993, 1999; Foody, 2002; Laba et al., 2002; Shao, Liu, & Zhao, 2001; Wickham, O'Neill, Ritters, Wade, & Jones, 1997; Yang, Stehman, Wickham, Smith, & Van Driel, 2000).

The standard approach for assessing classification accuracy is to select a sample of locations and determining the reference land cover present using field observations and/or fine resolution images. An error or confusion matrix is then formulated to catalog discrepancies between the land-cover map and the reference data (Congalton, Oderwald, & Mead, 1983; Story & Congalton, 1986). Various measures can then be derived from this table to report classification accuracy, including errors of omission and commission, producer's and user's accuracies and the Kappa coefficient (Congalton & Green, 1999).

A more thorough analysis of classification error would go beyond solely relying on the contingency table by incorporating contextual information, such as landscape characteristics in the analysis (Hubert-Moy, Cotonnec, Le Du, Chardin, & Perez, 2001; Pathirana, 1999; Shao et al., 2001; Sharma & Sarkar, 1998; Steele, Winne, & Redmond, 1998). Such analyses seek to provide additional insights into

* Corresponding author. Fax: +1-919-541-1138.
E-mail addresses: smith.jonathanh@epa.gov (J.H. Smith), svstehma@syr.edu (S.V. Stehman), wickham.james@epa.gov (J.D. Wickham), lyang@edcmail.cr.usgs.gov (L. Yang).
[1] Fax: +1-315-470-6535.
[2] Fax: +1-919-541-1138.
[3] Fax: +1-605-594-6529.

PII: S0034-4257(02)00126-8

Lichtenthaler, H. K., Lang, M., Sowinska, M., Heisel, F., & Miehe, J. A. (1996). Detection of vegetation stress via a new high resolution fluorescence imaging system. *J. Plant Physiol.*, *148*, 599–612.

Lichtenthaler, H. K., & Schweiger, J. (1998). Cell wall bound ferulic acid, the major substance of the blue-green fluorescence emission of plants. *J. Plant Physiol.*, *152*, 272–282.

Lopez-Hernandez, J., Vazquez-Oderiz, L., Vazquez-Blanco, E., Romero-Rodriguez, A., & Simal-Lozano, J. (1993). HPLC determinations of major pigments in the bean *Phaseolus vulgaris*. *J. Agric. Food Chem.*, *41*, 1613–1615.

McMurtrey III, J. E., Chappelle, E. W., Kim, M. S., Meisinger, J. J., & Corp, L. A. (1994). Distinguishing nitrogen fertilization levels in field corn (*Zea mays* L.) with actively induced fluorescence and passive reflectance measurements. *Remote Sens. Environ.*, *47*, 36–44.

Nilsson, H. E. (1995). Remote sensing and image analysis in plant pathology. *Ann. Rev. Phytopathol.*, *15*, 489–527.

Pinter Jr., P. J., Jackson, R. D., Ezra, C. E., & Gausman, H. W. (1985). Sun-angle and canopy-architecture effects on the spectral reflectance of six wheat cultivars. *Int. J. Remote Sensing*, *6*, 1813–1825.

Schowengerdt, R. A. (1997). *Remote sensing: models and methods for imaging processing* (p. 522). New York: Academic Press.

Schweiger, J., Lang, M., & Lichtenthaler, H. K. (1996). Differences in fluorescence excitation spectra of leaves between stressed and non-stressed plants. *J. Plant Physiol.*, *148*, 536–547.

Stober, F., & Lichtenthaler, H. K. (1993a). Characterization of the laser-induced blue, green, and red fluorescence signatures of wheat and soybean grown under different irradiances. *Physiol. Plant.*, *88*, 696–704.

Stober, F., & Lichtenthaler, H. K. (1993b). Studies on the constancy of the blue and green fluorescence yield during the chlorophyll fluorescence induction kinetics (Kautsky effect). *Radiat. Environ. Biophys.*, *32*, 357–365.

Stober, F. M., Lang, M., & Lichtenthaler, H. K. (1994). Blue, green, and red fluorescence emission signatures of green, etiolated, and white leaves. *Remote Sensing Environ.*, *47*, 65–71.

Valentini, R., Cecchi, G., Mazzinghi, P., Scarascia-Mugnozza, G., Agati, G., Bazzani, M., De Angelis, P., Fusi, F., Matteucci, G., & Raimond, V. (1994). Remote sensing of chlorophyll *a* fluorescence of vegetation canopies: 2. Physiological significance of fluorescence signal in response to environmental stresses. *Remote Sensing Environ.*, *47*, 29–35.

potential causes of classification error and possibly describe its spatial characteristics.

Patch size and land-cover heterogeneity are examples of landscape characteristics hypothesized to affect classification error (Campbell, 1996). Both result in an increase in the number of mixed pixels present in the land-cover data set. Mixed pixels record radiation reflectances arising from more than one land-cover class. Patch size and land-cover heterogeneity also influence classification error by introducing perceived pixel misclassifications when the land-cover map and reference data sets are misregistered, causing confusion as to the land cover actually present at a specific location.

The purpose of this article is to establish land-cover class-specific relationships between classification accuracy and two landscape variables, patch size and land-cover heterogeneity. Smith, Wickham, Stehman, and Yang (2002) investigated the effect of these two variables on classification error in general (i.e., not disaggregated by land-cover class). Building upon the previous finding that strong relationships do in fact exist, the current study focuses on evaluating these relationships for the individual land-cover classes. The results provide insight into how the accuracy of each land-cover class varies with changing landscape structure. Further, the logistic regression relationships established for each class provide a useful, qualitative predictive tool for determining where classification error is most likely to occur. The goal is to better illuminate the link between accuracy information, the land-cover map and the landscape being studied.

2. Methodology

The 1990 National Land-Cover Data (NLCD) set covers the conterminous United States at a pixel resolution of 30 m (Vogelmann, Howard, Yang, Larson, Wylie, & Van Driel, 2001). Produced by the Multi-Resolution Land Characteristics (MRLC) consortium (Loveland & Shaw, 1996), the NLCD was derived from Landsat Thematic Mapper (TM) images (Vogelmann, Sohl, & Howard, 1998; Vogelmann et al., 2001). The NLCD employs a land-cover classification scheme modeled upon the Anderson, Hardy, Roach, and Witmer (1976) system at two classification levels (Table 1).

Accuracy assessment of the NLCD set is being implemented by EPA federal region, with reference sample data utilized in this study encompassing regions 1, 2, 3 and 4 (Fig. 1). As a result of the regional progression of accuracy assessments, different photo-interpreter teams acquired the reference data in each of these four regions. Accuracy assessment methodology was based on a probability sample of pixels, with reference land-cover data obtained by interpretation of hard-copy, National Aerial Photography Program (NAPP) photographs (Stehman, Wickham, Yang, & Smith, 2000; Yang, Stehman, Smith, & Wickham, 2001; Yang et al., 2000; Zhu, Yang, Stehman, & Czaplewski, 1999; Zhu, Yang, Stehman, & Czaplewski, 2000). A sample pixel was considered correctly classified if the primary photo-interpreted class matched the NLCD class. Other definitions of agreement have been employed in the reporting of NLCD accuracy results (Yang et al., 2001; Zhu et al., 1999).

Two landscape variables, land-cover heterogeneity and patch size, were recorded for each sample pixel. Values for

Table 1
Major 1992 national land-cover classes found in regions 1–4

Level I	Level II	Class	Definition
20	21	Low Intensity Residential	Mixture of constructed materials and vegetation, with the constructed materials accounting for 30–79% of the cover.
	22	High Intensity Residential	Mixture of constructed materials and vegetation, with the constructed materials accounting for 80–100% of the cover.
	23	Commercial/ Industrial/ Transportation	Highly developed areas not classified as high intensity residential.
30	31	Bare Rock/ Sand/Clay	Perennially barren areas of earthen materials.
	32	Quarries/ Strip Mines/ Gravel Pits	Areas of extractive mining activities with significant surface expression.
	33	Transitional	Areas of sparse vegetation cover (< 25%) that are dynamically changing from one land cover to another.
40	41	Deciduous Forest	Areas dominated by trees in which >75% of the trees shed foliage spontaneously in response to seasonal changes.
	42	Evergreen Forest	Areas dominated by trees in which >75% of the trees retain foliage all year.
	43	Mixed Forest	Areas inhabited by both deciduous and evergreen trees with neither comprising >75% of total tree cover.
80	81	Pasture/Hay	Areas of grasses, legumes, or grass–legume mixtures planted for livestock grazing, or the production of seed, or hay crops.
	82	Row Crops	Areas used for the production of crops such as corn soybeans, vegetables, tobacco and cotton.
	85	Urban/Recreational Grasses	Vegetated areas in developed settings set aside for the purposes of recreation, erosion control, or aesthetics.
90	91	Woody Wetlands	Areas with forest, or shrubs accounting for 25–100% of the cover and periodically saturated with water.
	92	Emergent Herbaceous Wetlands	Areas with perennial herbaceous vegetation accounting for 25–100% of the cover and periodically saturated with water.

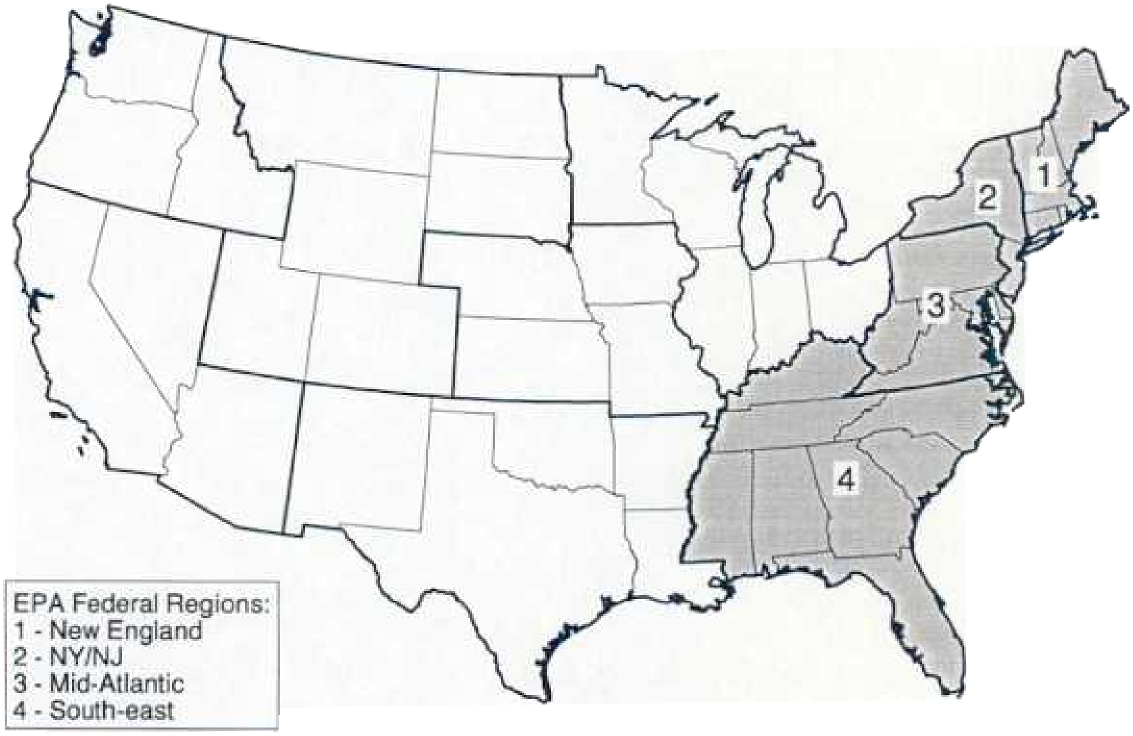

Fig. 1. Area of classification accuracy

each variable were obtained for the Level II classification scheme, which is the basic scheme of the NLCD and the more generalized Level I scheme (see Table 1). Thus, each sample point had two heterogeneity and two patch size variable values (one at Level I and another at Level II).

Land-cover heterogeneity was computed to equal the number of land-cover classes occurring in a 3×3 pixel window centered on the sample pixel. Heterogeneity values ranged from 1, which meant that a single land-cover class occurred in the window, to 7 for the Level II classification scheme and from 1 to 6 for the Level I scheme. A heterogeneity value of 1 indicates that the sample pixel is located within a homogeneous 3×3 block of pixels (an interior pixel), while any value greater than 1 indicates that the pixel was located on a patch edge. In place of this heterogeneity variable, a simpler dichotomous variable indicating whether the sample pixel was an interior or edge pixel was also considered. Based on logistic regression analyses not reported in this article, we found that this dichotomous heterogeneity variable contributed less to the explanatory ability of the logistic regression models than did the quantitative heterogeneity variable for all Levels I and II classes except for bare rock (31). Consequently, we retained the heterogeneity variable for all subsequent analyses.

The other landscape variable, patch size, was calculated to equal the number of contiguous pixels of the same land-cover class. Contiguity was defined as occurring when two pixels of the same class were adjacent, including diagonally, to one another. A buffering operation was implemented to efficiently process the land-cover data for patch size calculation and also to impose an upper bound on patch size. Buffers with a radius of 3000 m were created around the accuracy assessment sample points and then overlayed upon both the Levels I and II land-cover data sets, resulting in circular zones of land cover 200 pixels across. Patch sizes ranged from 1 to 139,317 pixels. Preliminary analysis of this data indicated that a transformation to a logarithmic (base 10) scale would improve the linearity of the logistic regression models. Any further reference to the patch size variable should be understood as pertaining to the logarithm of patch size. The characteristics of both landscape variables are summarized for each of the land-cover classes in Table 2.

In addition to the two landscape variables, three regional dummy variables were also created and treated as a set in the statistical analyses. These variables account for potential regional differences in physiography, photo-interpretation skills and image classification protocols. Additionally, they were included to assess whether the effect of the two landscape variables could be attributed to their confounding with regional differences. The final explanatory variable was an interaction term calculated to equal the product of the heterogeneity and patch size variables. It was included to assess whether the two landscape variables interact to influence classification accuracy.

The dichotomous response variables recorded for each sample pixel represented whether the pixel was correctly classified or not. Separate response variables were calculated for each classification level. Response variables were coded as 1 if the pixel was correctly classified and 0 if it was misclassified. Logistic regression was then used to

Table 2
Description of NLCD land cover at the sample points

	Class	Number of samples	Percent correctly classified	Average heterogeneity	Standard deviation heterogeneity	Average patch size	Standard deviation patch size	Correlation of landscape variables
Level II	21	323	38	2.57	1.11	2.65	1.37	−0.503
	22	332	28	2.20	0.98	2.55	1.16	−0.505
	23	320	48	2.68	1.24	2.18	1.20	−0.534
	31	291	33	2.56	1.22	1.90	1.02	−0.575
	32	310	34	2.09	1.25	2.42	1.00	−0.603
	33	305	37	2.25	1.13	2.12	1.08	−0.564
	41	630	48	1.86	0.93	3.47	1.40	−0.564
	42	337	45	2.22	0.99	2.52	1.26	−0.460
	43	449	33	2.41	0.87	2.34	1.32	−0.503
	81	447	33	2.10	1.09	2.78	1.27	−0.593
	82	326	40	2.17	1.10	2.48	1.19	−0.486
	85	326	41	2.61	1.21	1.76	1.03	−0.592
	91	305	33	2.22	1.19	2.59	1.39	−0.629
	92	319	54	2.26	1.35	2.62	1.65	−0.743
Level I	20	975	70	1.74	0.82	3.76	1.27	−0.484
	30	906	43	2.07	0.92	2.19	1.00	−0.536
	40	1416	77	1.33	0.56	4.40	0.95	−0.466
	80	1099	50	1.76	0.79	3.08	1.35	−0.551
	90	624	55	1.82	0.84	2.86	1.50	−0.597

evaluate relationships between these response variables and various sets of explanatory variables (i.e., the landscape, interaction and regional variables). Rather than model directly the dichotomous response variable, logistic regression instead models the logarithm of the odds, where the odds of a correct classification is defined as $p/(1-p)$, with p being the probability of a correct classification. The logistic regression model is:

$$\ell n(p/(1-p)) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \quad + \beta_k x_k \qquad (1)$$

where α is the intercept, x_1 through x_k are explanatory variables, β_1 through β_k are the parameters and ℓn is the natural logarithm. The model assumes that the response variable represents the outcome of a Bernoulli trial and that responses of different sample elements are independent.

Several logistic regression models (Table 3) were evaluated to investigate various features of the explanatory variables' effects. These models possess a hierarchical structure such that a model with one or few explanatory variables (i.e. a reduced model) nests within a more encompassing model (i.e. a full model) containing the reduced model's explanatory variables plus one or more additional variables. Statistical tests were then conducted to determine if the additional explanatory variable or variables present in the full, but not reduced, model contributed a significant improvement in model fit. That is, the tests evaluated the marginal contribution of these explanatory variables to a model already containing other explanatory variables. The test employs a chi-square statistic derived from differences in the -2 log likelihood (-2LL) values for the full and reduced models (Agresti, 1996; Hosmer & Lemeshow, 1989). Statistical significance was judged based on a significance (α) level of 0.05.

In the results presented for the single-variable logistic regression models (Models 1a and 1b), we rely heavily on the correspondence between the coefficient β_1 and the odds ratio. An odds ratio is defined as:

$$(p_1/(1-p_1)/p_2/(1-p_2)) \qquad (2)$$

where p_1 and p_2 are the probabilities of a correct classification at two different levels of explanatory variable x. Odds ratios provide a convenient metric for assessing the relative change in the odds of a correct classification given a one unit change in x. An odds ratio of 1 indicates that no change in the odds of a correct classification is associated with a one unit change in the explanatory variable. Equivalently, an odds ratio of 1 occurs when $\beta_1 = 0$, a situation in which the odds of a correct classification shows no linear relationship with the explanatory variable. An odds ratio greater than 1 indicates that the odds of a correct classification increases as the

Table 3
Logistic regression models evaluated

Model number	Model	Description
0	β_0	
1a	$\beta_0 + \beta_1 x_1$	
	$\beta_0 + \beta_2 x_2$	
	$\beta_0 + \beta_1 x_1 + \beta_2 x_2$	
3	$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{1,2} x_1 x_2$	
	$\beta_0 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$	
	$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$	

variable increases by one unit. Conversely, an odds ratio less than 1 indicates a decrease in the odds of a correct classification when the variable value increases. The magnitude of the deviation of the odds ratio from 1 represents the "strength" of the change in the odds as the explanatory variable increases by one unit. Alternatively, the deviation of the estimated odds ratio from 1 may be interpreted as representing the "sensitivity" of that land-cover class to changes in the landscape variable.

3. Results

For single-variable Models 1a and 1b at classification Level II, both landscape variables were statistically significant for all but a few land-cover classes. Land-cover heterogeneity was not significant in the low intensity residential (21) and mixed forest (43) classes, while patch size was not significant in the pasture/hay (81) class. The general impacts of the two variables were consistent. Odds of a correct classification increased as heterogeneity decreased and as patch size increased (Table 4). However, the estimated coefficients and, accordingly, the estimated odds ratios vary broadly among the land-cover classes for both landscape variables. This diversity of effects illustrates the land-cover class specificity of the relationships between classification accuracy and the landscape variables.

Table 4
Results for single-variable logistic regression models (1a and 1b) at Level II

Explanatory variable	Class	Parameter estimates		95% profile likelihood confidence limits of the odds ratio		
		Intercept α	Slope β	Lower	Estimate	Upper
Land-cover heterogeneity	21	−0.30	−0.07	0.76		
	22	0.09	−0.49	0.46		
	23	1.36	−0.55	0.47		
	31	0.30	−0.42	0.52		
	32	0.60	−0.67	0.39		
	33	0.02	−0.26	0.62		
	41	1.34	−0.77	0.38		
	42	1.32	−0.70	0.38		
	43	−0.52	−0.07	0.74		
	81	0.12	−0.35	0.57		
	82	0.74	−0.57	0.46		
	85	0.58	−0.37	0.56		
	91	0.18	−0.41	0.53		
	92	2.62	−1.14	0.25		
Patch size	21	−1.08	0.22	1.06		
	22	−2.44	0.55	1.37		
	23	−1.58	0.67	1.58		
	31	−2.54	0.89	1.80		
	32	−2.48	0.71	1.50		
	33	−1.45	0.41	1.20		
	41	−1.76	0.48	1.42		
	42	−0.86	0.26	1.09		
	43	−1.17	0.20	1.05		
	81	−0.78	0.07	0.90		
	82	−1.63	0.45	1.31		
	85	−1.10	0.40	1.19		
	91	−1.65	0.36	1.19		
	92	−3.22	1.28	2.82		

Table 5
Results for single-variable logistic regression models (1a and 1b) at Level I

Explanatory variable	Class	Parameter estimates		95% profile likelihood confidence limits of the odds ratio		
		Intercept α	Slope β	Lower	Estimate	Upper
Land-cover heterogeneity	20	2.4	−0.86	0.35		
	30	0.95	−0.60	0.47	0.55	0.64
	40	3.14	−1.38	0.20	0.25	0.31
	80	2.21	−1.28	0.23	0.28	0.34
	90	1.88	−0.93	0.31	0.39	0.49
Patch size	20	−1.71	0.70	1.78	2.01	2.27
	30	−1.62	0.60	1.58	1.83	2.13
	40	−1.70	0.67	1.72	1.95	2.23
	80	−2.29	0.73	1.86	2.08	2.33
	90	−1.88	0.73	1.82	2.07	2.37

To illustrate the interpretation of the odds ratios for land-cover heterogeneity, we focus on two Level II classes, emergent herbaceous wetland (92) and transitional (33). For the heterogeneity variable, emergent herbaceous wetland had the smallest odds ratio, 0.32 (i.e. it was the most sensitive to heterogeneity changes), while the transitional class had the largest, 0.77 (i.e. it was the least sensitive). Therefore, the odds of correctly classifying an emergent herbaceous wetland pixel having a heterogeneity value of 2 would be 3.1 (1/0.32) times lower than the odds of correctly classifying an interior (heterogeneity value of 1) emergent herbaceous wetland pixel. In contrast, the odds of correctly classifying a transitional class pixel is only 1.3 (1/0.77) times lower for the same change in heterogeneity.

Emergent herbaceous wetland was also the most sensitive to changes in patch size, with an estimated odds ratio of 3.6. Accordingly, as patch size increases from 100 to 1000 pixels (an increase in the logarithm value of 2 to 3), the odds of the pixel in the larger patch being correctly classified was 3.6 times higher than the pixel in the smaller patch. The class least impacted by changes in patch size was the commercial/industrial/transportation class (23), which had an odds ratio of 1.22.

The single-variable model results (Table 5) for Level I follow a pattern similar to the Level II results. Both land-cover heterogeneity and patch size were significant for all five Level I classes and the expected relationships that the odds of correct classification increase as patch size increases and as heterogeneity decreases were found. At Level I, forest (40) was most sensitive to heterogeneity, while the barren class (30) was least sensitive to this variable. The barren class also showed the least sensitivity to patch size, while agriculture (80) was the most sensitive.

Relative importance of the two landscape variables was evaluated by comparing Model 2 with Models 1a and 1b. Because of the correlation between the two landscape variables (see Table 2), the relative importance of each variable must be assessed by testing its marginal contribution adjusted for explanatory ability shared with the other variable (Table 6). The marginal contribution of heterogeneity adjusted for patch size was not statistically significant for the following Level II classes: low density residential (21), high density residential (22), bare rock (31), transitional (33), mixed forest (43) and both wetland classes (91, 92). For these same classes, the marginal contribution of patch size was statistically significant when adjusted for the presence of heterogeneity in the model. This suggests that patch size may be the more important of the two explanatory variables for these classes and a single-variable model using only patch size is as good a model as the two-variable model. For evergreen forest (42), pasture/hay (81) and urban/recreational grass (85), heterogeneity appeared to be the more important variable, since it remained significant in the presence of patch size, but the marginal contribution of patch size adjusted for heterogeneity was not statistically significant. For the remaining Level II classes, commercial/industrial (23), quarries/strip mines (32), deciduous forest (41), and row crops (82), the marginal contributions of both landscape variables were statistically significant when adjusted for the influence of the other variable. This characteristic was also the case for all five Level I classes: both landscape variables were needed in the model.

Comparing Model 2 with Model 3 evaluates whether the patch size by land-cover heterogeneity interaction term was required in the model. For all Level II classes, this term was not statistically significant, while at Level I, it was significant for only the urban class (20). The absence of a significant interaction effect suggests that the effect of heterogeneity remains the same regardless of the value of patch size and vice versa. For the Level I urban class, the effect of heterogeneity would vary depending on the size of the patch in which the sample pixel is located.

The final two model comparisons evaluate the importance of the set of regional dummy variables relative to the two landscape variables. For all classes at both classification levels, the marginal contribution of the landscape variables was statistically significant when adjusted for the effect shared by the regional variables (Model 4 vs. Model 5). Consequently, the effects of the landscape variables cannot be dismissed as resulting from a confounding effect attributable to their association with regional variation. Conversely, the set of regional dummy variables did not provide a significant marginal contribution when adjusted for the landscape variables (Model 2 versus Model 5) in four of the Level II classes, transitional (33), deciduous forest (41) and both wetland classes (91, 92) and two of the Level I classes, urban (20) and forest (40). For these classes, no additional explanatory ability pertaining to accuracy is obtained from the regional variables once they have been adjusted for the landscape variables. The regional variables were statistically important for all of the other classes, suggesting that some of the variability in classification accuracy not explainable by the landscape variables can be attributed to characteristics associated with the different regions.

Table 6
Chi – square analysis of model comparisons

Model comparison	Degrees of freedom	Level II land-cover class													
		21	22	23	31	32	33	41	42	43	81	82	85	91	92
Model 0–Model 2	2	7.56	25.82	51.32	41.49	37.48	13.27	83.08	33.31	7.57	13.96	37.89	17.51	17.89	195.22
Model 1a–Model 2	1	ns	ns	6.89	ns	11.79	ns	21.81	24.66	ns	13.34	10.99	4.57	ns	ns
Model 1b–Model 2	1	7.14	13.04	19.52	27.08	6.62	7.64	17.36	ns	7.22	ns	9.19	ns	4.18	82.13
Model 2–Model 3	1	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Model 4–Model 5	2	10.89	12.38	54.71	29.76	41.46	14.73	83.51	32.37	12.28	19.79	27.44	7.52	14.52	150.68
Model 2–Model 5	4	22.74	20.58	21.38	37.87	20.4	ns	ns	38.45	68.57	20.31	98.95	14.31	ns	ns

Model comparison	Degrees of freedom	Level I land-cover class				
		20	30	40	80	90
Model 0–Model 2	2	177.51	86.53	199.29	275.73	153.23
Model 1a–Model 2	1	21.82	15.14	80.55	66.7	5.03
Model 1b–Model 2	1	79.02	27.15	32.07	68.45	74.31
Model 2–Model 3	1	12.99	ns	ns	ns	ns
Model 4–Model 5	2	170.84	80.5	194.11	282.24	135.9
Model 2–Model 5	4	ns	15.57	ns	17.39	9.82

ns—not significant at $\alpha = 0.05$ significant level.

Explanation of difference tests: (a) Model 0–Model 2: Is the joint contribution of patch size and heterogeneity significant? (b) Model 1b–Model 2: Is the additional explanatory contribution of patch size to a model already containing heterogeneity significant? (c) Model 1a–Model 2: Is the additional explanatory contribution of heterogeneity to a model already containing patch size significant? (d) Model 2–Model 3: Does the interaction between patch size and heterogeneity contribute significant explanatory value? (e) Model 4–Model 5: Is the additional explanatory contribution of the landscape variables to a model already containing the regional dummy variables statistically significant? (f) Model 2–Model 5: Is the additional explanatory contribution of the regional variables to a model already containing the landscape variables statistically significant?

4. Conclusions

The goal of this paper was to analyze the class-specific impacts of landscape characteristics on classification accuracy. Analyses were performed using 5020 photo-interpreted assessment points, across 21 states, at two classification levels. Land-cover heterogeneity and patch size were found to be important factors determining land-cover classification accuracy, with the general effects of the variables holding steady across classes: odds of a correct classification increases with increasing patch size and decreasing heterogeneity. However, the land-cover classes display marked individuality in the specific nature of these relationships. Some classes demonstrate greater sensitivity to heterogeneity, others to patch size, with still others strongly affected by both variables. Modeling the effects of landscape structure on classification error clearly should be pursued on an individual land-cover class basis.

For all classes at both Levels I and II, the landscape variables maintained their importance when adjusted for regional effects as represented by the set of regional dummy variables. However, for most of the land-cover classes, the best models were those in which both the regional variables and one or both landscape variables were included. The patch size by heterogeneity interaction term did not meaningfully contribute to the explanation of classification accuracy (except for Level I urban). That is, the effect of heterogeneity on accuracy is the same whether the pixel was found in a large or small patch. This suggests a localized spatial effect in that if the pixel was in a heterogeneous area (i.e., near one or more edges), it does not matter if the area of the homogeneous land-cover patch within which the pixel fell was large or small.

Quantifying the relationships by land-cover class offers useful descriptive information regarding the nature of classification errors in the NLCD maps. For any specific map location, we can readily observe the land-cover class and visually determine the local landscape structure. Based on the class-specific models and the estimated parameters of patch size and land-cover heterogeneity (see Table 4), we can approximate the odds of that location being correctly classified. For example, if an emergent herbaceous wetland (92) pixel is found in a small patch in a very heterogeneous locale, the odds are high that the pixel will be misclassified relative to the odds of a pixel in a larger patch with less heterogeneous land-cover. Conversely, a pixel classified as transitional (33) found in a small patch in a heterogeneous locale will not differ as much in its odds of a correct classification as compared to a pixel in a larger patch within more homogeneous land cover. Practically speaking, heterogeneity surrounding an emergent herbaceous wetland pixel raises a caution that classification error is much more likely, whereas heterogeneity surrounding a transitional class pixel causes not as much concern.

A goal of NLCD accuracy assessment research is to derive land-cover class-specific predictive models of classification error using readily available explanatory variables such as landscape structure. Such models may serve as a simple method for creating maps qualitatively representative of classification error in the NLCD. The results reported in this study not only provide insight into factors associated with classification error, they also serve as an important intermediate step in the development of a predictive capacity for modeling error.

Acknowledgements

This work was supported by the US Environmental Protection Agency and has been approved for publication. Dr. Stephen Stehman's work was performed under US Environmental Protection Agency contract 0V1036NTSX. Dr. Limin Yang's work was performed under US Geological Survey contract 1434-CR-97-CN-40274. The authors would like to thank Dr. Brian Steele and an anonymous reviewer for their helpful comments. Mention of trade names or commercial products does not constitute an endorsement or recommendation for use.

References

Agresti, A. (1996). *An introduction to categorical data analysis*. New York, NY: Wiley, 290 pp.

Anderson, J. F., Hardy, E. E., Roach, J. T., & Witmer, R. E. (1976). A land use and land cover classification system for use with remote sensor data. *U.S. Geological Survey Professional Paper 964*. Washington, DC: U.S. Geological Survey, 28 pp.

Campbell, J. B. (1996). *Introduction to remote sensing* (2nd ed.). New York, NY: Guilford Press, 622 pp.

Congalton, R. G., & Green, K. (1993). A practical look at the sources of confusion in error matrix generation. *Photogrammetric Engineering and Remote Sensing*, *59*(5), 641–644.

Congalton, R. G., & Green, K. (1999). *Assessing the accuracy of remotely sensed data: principles and practices*. Boca Raton, FL: Lewis Publishers, 137 pp.

Congalton, R. G., Oderwald, R. G., & Mead, R. A. (1983). Assessing Landsat classification accuracy using discrete multivariate statistical techniques. *Photogrammetric Engineering and Remote Sensing*, *49*(12), 1671–1678.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, *80*, 185–201.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York, NY: Wiley, 307 pp.

Hubert-Moy, L., Cotonnec, A., Le Du, L., Chardin, A., & Perez, P. (2001). A comparison of parametric classification procedures of remotely sensed data applied on different landscape units. *Remote Sensing of Environment*, *75*, 174–187.

Jensen, J. R. (1996). *Introductory digital image processing, a remote sensing perspective* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall, 316 pp.

Jensen, J. R. (2000). *Remote sensing of the environment: an earth resource perspective* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall, 544 pp.

Laba, M., Gregory, G. J., Ogurcak, D., Hill, E., Fegraus, E., Fiore, J., & DeGloria, S. D. (2002). Conventional and fuzzy accuracy assessment of the New York gap analysis project land cover map. *Remote Sensing of Environment*, *81*, 443–455.

Loveland, T. R., & Shaw, D. M. (1996). Multiresolution land character-